Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays

Fengli Fu[1,2], Jeffry D. Sander[1,2], Morgan Maeder[3], Stacey Thibodeau-Beganny[3], J. Keith Joung[3,4], Drena Dobbs[1,2], Leslie Miller[2,5] and Daniel F. Voytas[6,*]

[1]Department of Genetics, Development & Cell Biology, 1043 Roy J. Carver Co-Laboratory, [2]Interdepartmental Graduate Program in Bioinformatics & Computational Biology, 2114 Molecular Biology Building, Iowa State University, Ames, Iowa 50011, [3]Molecular Pathology Unit, Center for Cancer Research and Center for Computational and Integrative Biology, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, [4]Department of Pathology, Harvard Medical School, Boston, MA 02115, [5]Department of Computer Science, 227 Atanasoff Hall, Iowa State University, Ames, IA 50011 and [6]Department of Genetics, Cell Biology & Development and Center for Genome Engineering, 321 Church Street SE, University of Minnesota, MN 55455, USA

ABSTRACT

Zinc fingers are the most abundant DNA-binding motifs encoded by eukaryotic genomes and one of the best understood DNA-recognition domains. Each zinc finger typically binds a 3-nt target sequence, and it is possible to engineer zinc-finger arrays (ZFAs) that recognize extended DNA sequences by linking together individual zinc fingers. Engineered zinc-finger proteins have proven to be valuable tools for gene regulation and genome modification because they target specific sites in a genome. Here we describe ZiFDB (Zinc Finger Database; http://bindr.gdcb.iastate.edu/ZiFDB), a web-accessible resource that compiles information on individual zinc fingers and engineered ZFAs. To enhance its utility, ZiFDB is linked to the output from ZiFiT—a software package that assists biologists in finding sites within target genes for engineering zinc-finger proteins. For many molecular biologists, ZiFDB will be particularly valuable for determining if a given ZFA (or portion thereof) has previously been constructed and whether or not it has the requisite DNA-binding activity for their experiments. ZiFDB will also be a valuable resource for those scientists interested in better understanding how zinc-finger proteins recognize target DNA.

INTRODUCTION

The C2H2 zinc-finger motif is approximately 30 amino acids in length and conforms to the pattern (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X is an amino acid and ψ is a hydrophobic residue (1). The amino acids fold into a $\beta\beta\alpha$ structure, which is stabilized by a zinc ion coordinated by the two conserved cysteine and histidine residues. In binding DNA, each finger typically recognizes three adjacent nucleotides: amino acids at positions -1, $+3$ and $+6$ (relative to the beginning of the α-helix) contact nucleotides at the 3′-end, middle and 5′-end of the target nucleotide triplet. Zinc fingers can function as monomers, making it possible to link zinc fingers together in extended arrays that recognize unique DNA sequences. Such artificial or engineered zinc-finger arrays (ZFAs) are often fused to effector domains (e.g. transcriptional activation domains or nucleases), and engineered zinc-finger fusion proteins are proving increasingly valuable as reagents for gene regulation and genome modification (2).

A simple approach for engineering ZFAs is known as 'modular assembly', wherein individual fingers are joined together to create an array that recognizes a novel target sequence. To facilitate modular assembly, efforts have been made to obtain zinc fingers that recognize all 64 possible triplets. Approaches to create such a finger archive have included parallel selection from random libraries (3–7), rational design (8,9) and characterization of naturally occurring zinc-finger proteins (10). Zinc fingers have thus far been identified that recognize all 16 GNN triplets as well as most CNN, ANN and some TNN triplets. Modular assembly assumes that individual zinc fingers are not influenced by neighboring fingers in an array. However, systematic testing of a large number of ZFAs constructed with available zinc-finger modules resulted in success rates of 56%, 20%, 4% and 0% for

*To whom correspondence should be addressed. Tel: +1 515 294 1963; Fax: +1 515 294 7155; Email: voytas@iastate.edu

9-bp target sites composed of three, two, one or no GNN triplets, respectively (11). It is clear from this work that zinc fingers are not always modular and that a better understanding of context-dependent effects is needed to increase the reliability of modular assembly.

To minimize the context dependence associated with modular design, sequential selection (12), bipartite selection (13) and bacterial-based selection methods have been described (14,15). These methods typically involve screening large libraries of ZFA variants. The few arrays in the library that recognize a specific target with high affinity and specificity are identified using phage display or a bacterial two-hybrid system in which target binding activates expression of a reporter gene. An advantage of these selection-based strategies is that context-dependent effects are addressed through the process of selection. Selection-based approaches, however, require considerable molecular biological expertise and are time consuming. Recently, a selection platform was described, called OPEN (Oligomerized Pool ENgineering), that overcomes shortcomings of traditional selection-based methods in that it is simpler and faster; it is also much more effective than modular assembly (11,16). However, even the simplified OPEN strategy remains more labor-intensive and difficult to perform than modular assembly.

Because many scientists are interested in generating zinc-finger proteins for gene regulation and genome modification, we implemented a Zinc Finger Database (ZiFDB)—an archive of information about zinc fingers and engineered ZFAs. The database is one of several, integrated resources provided by the Zinc Finger Consortium (http://www.zincfingers.org), a group of academic laboratories dedicated to improving methods for engineering ZFAs and developing new applications for their use. The Consortium has previously created a repository of zinc-finger encoding plasmids which simplify and facilitate modular assembly of ZFAs (17). Consortium labs also developed and validated the OPEN platform for ZFA engineering (11,16). Further, Consortium-supported software, called ZiFiT (Zinc Finger Targeter), allows users to scan DNA sequences to identify potential sites for which ZFAs might be engineered by either modular assembly or OPEN (18). A particularly important feature of ZiFDB is its linkage to ZiFiT, which allows users to readily determine if particular ZFAs have been previously made by other investigators that recognize a target sequence of interest.

DATABASE CONTENTS

Zinc fingers

ZiFDB currently contains 716 zinc fingers, each of which is assigned a unique numerical designator. The database was initially populated with fingers generated and characterized by four different research groups:

(1) Sangamo BioSciences: Sangamo Biosciences designed zinc fingers that recognize GNN triplets at each of the three positions in a three-finger array (7). It is assumed that these fingers function optimally in the position for which they were designed. Sangamo has also generated several zinc fingers that recognize non-GNN triplets, and these are also included in the database (17,19,20).

(2) Barbas laboratory: The Barbas laboratory at the Scripps Research Institute developed a large number of zinc fingers that recognize diverse nucleotide triplets. The fingers are presumed to be modular and function at any position within a ZFA (3,4,6,21).

(3) Toolgen: Toolgen Inc. assembled a collection of zinc fingers derived from naturally occurring zinc-finger proteins encoded in the human genome (10).

(4) Joung laboratory: The Joung laboratory at Massachusetts General Hospital used the OPEN platform to generate a large number of ZFAs (16). All of these ZFAs are three finger domains that recognize a 9-bp target. Unique zinc fingers comprising ZFAs generated by OPEN are also included in the database.

In some cases, investigators have used the DNA-recognition helices from the above fingers within the context of a different zinc-finger backbone. In ZiFDB, such fingers are given a different numerical designator, since the backbone may affect zinc-finger function.

The following information is provided for each zinc finger:

(1) Triplet target: The sequence is provided for the cognate 3-bp DNA target of each finger.

(2) Recognition helix: The sequence is provided for the seven amino acids that constitute the recognition helix: positions −1 to +6 relative to the start of the α-helix that contacts DNA in the finger.

(3) Position within a three-finger array: As indicated above, some fingers were designed to function optimally at certain positions within a three-finger array. The fingers described by Sangamo BioSciences, for example, were selected at different positions within a three-finger protein, and the recognition helices for the same triplet target at different positions in a three finger protein usually differ (7). These differences are likely due to position and context-dependent effects. In addition, the fingers generated by OPEN were selected to function in a specific position within a three finger protein (16). The database records position information for the fingers described by Sangamo BioSciences and generated by OPEN. By convention, finger 1 (F1) specifies the 3′-triplet in a 9-bp target, F2 specifies the middle triplet and F3 specifies the 5′-triplet.

(4) Source: The name of the group that generated the zinc finger is provided. The groups at present include the Barbas and Joung laboratories, Sangamo BioSciences (SGMO) and ToolGen.

(5) Amino acid sequence: The amino acid sequence of the entire finger is provided.

(6) Article: A link to the relevant citation is provided so the user can get more detailed information about a particular finger.

(7) Experiment: A link directs the user to information describing the method used to generate the finger, relevant functional data and other observations made by the experimenter.

Zinc-finger arrays

At present, ZiFDB houses 652 three-finger arrays collected from the published literature. Although engineered ZFAs of four or more fingers have been reported, the engineering platforms for which the Zinc Finger Consortium has developed reagents are only designed for constructing three-finger arrays. Larger arrays, therefore, are currently not supported by ZiFDB. Each three-finger array in the database has a unique numerical designator, and the following information is provided:

(1) Identification numbers (IDs) of the component fingers: The IDs correspond to the zinc-finger designator and are hyperlinks that allow a user to look up detailed information about individual fingers in an array.
(2) Binding subsites: These sequences correspond to the 3-bp nucleotide target of the corresponding component finger.
(3) Recognition helix: For each finger in an array, the sequence is provided of the seven amino acids that make up the recognition helix.

(4) Article: A link to the relevant citation is provided so the user can get more detailed information about that array.
(5) Experiment: A link is provided to information describing the method used to generate the arrays and relevant functional data or other observations made by the experimenter.

Database interface

To retrieve information about a particular zinc finger, any combination of three search criteria can be used: triplet target, finger position and/or finger source. By providing partial information, a list of all fingers matching the input is returned (Figure 1). By clicking the link in the article or experiment column, detailed information about related citations or experiments is displayed.

When searching for information about an array, the user provides the nucleotide triplets (subsites) recognized by each finger in the array. The database returns not only the arrays recognizing the three input subsites, but also arrays matching any two of three triplet subsites (Figure 2). This partial information may inform the design of novel ZFAs by providing the user with a two finger protein that could be modified and tested for function. For arrays that match only two of the input subsites, the matching subsites and their recognition helices are colored

IOWA STATE UNIVERSITY

ZiFDB version 1.0

Department of Genetics, Development and Cell Biology
Interdepartmental Graduate Program in Bioinformatics and Computational Biology

Menu

Introduction

Instructions

Search ZiFDB

Insert article

Links

ZiFiT software

Zinc Finger Consortium

Addgene

Your search results:

ID	Target	Recognition helix	Position	Source	Sequence	Article ID	Experiment ID
11	GGG	RSDHLAR	F1	SGMO	PGEKPHICHIQGCGKVYGRSDHLARHLRWH	1	
27	GGG	RSDHLTR	F2	SGMO	PGERPFMCTWSYCGKRFTRSDHLTRHKRTH	1	
51	GGG	RSDHLSR	F3	SGMO	PGEKKFACPECPKRFMRSDHLSRHIKTH	1	
58	GGG	RSDKLVR	F2	Barbas	PGEKPYKCPECGKSFSRSDKLVRHQRTH	8	
131	GGG	RKDHLTR	N/A	ToolGen	PGEKPFLCQYCAQRFGRKDHLTRHMKKSH	12	
133	GGG	RSDKLNR	N/A	ToolGen	PGEKPYVCDVEGCTWKFARSDKLNRHKKRH	12	
134	GGG	RRSHLTR	N/A	ToolGen	PGEKPYKCMECGKAFNRRSHLTRHQRIH	12	
298	GGG	RSAHLQA	F1	Joung	PHERPFQCRICMRNFSRSAHLQAHTRTH	19	
299	GGG	RGEHLTN	F1	Joung	PHERPFQCRICMRNFSRGEHLTNHTRTH	19	
300	GGG	RRLHLTN	F1	Joung	PHERPFQCRICMRNFSRRLHLTNHTRTH	19	
301	GGG	KQDHLTK	F1	Joung	PHERPFQCRICMRNFSKQDHLTKHTRTH	19	

Voytas Laboratory
Current address: University of Minnesota
Phone:(612) 626-4509
Email: voytas@umn.edu

Figure 1. Sample output from the finger search page when GGG is provided as the target and the position and source are left blank.

Figure 2. Sample output returned when searching for arrays with the binding site 5′-GAAGGCGGC-3′.

in red for easy identification. The output also provides the amino acid sequence of the array, IDs for each finger in the array and IDs for related articles or experiments. By clicking the relevant ID, the user is directed to information on specific fingers, articles or experiments.

Interface with pre-existing ZiFiT software

To enhance its utility, ZiFDB is interfaced directly with ZiFiT—a web accessible software package that assists biologists with zinc-finger protein design (http://bindr.gdcb. iastate.edu/ZiFiT/) (16,18). ZiFiT provides users with a list of target sites within their DNA sequence of interest as well as the corresponding Consortium-developed reagents available for engineering the ZFAs. Target site hyperlinks within the ZiFiT output directly query ZiFDB to determine if any previously constructed arrays exist that bind to completely or partially matched target sequences. In addition, ZiFiT users can query ZiFDB for finger information for a specific triplet subsite by clicking on the triplet. Thus, ZiFiT and ZiFDB work synergistically to aid in ZFA design.

Implementation

ZiFDB stores zinc-finger information as a set of objects defined by Java classes. The atomic information is contained in two main classes (Zinc Finger and Zinc Finger Array). Both classes reference additional Article and Experiment classes. The Article class includes the title, journal, volume, page, year and authors for the article being represented. The Experiment class includes information on the submitter, assay, results and other comments provided by the author or submitter. The authors of articles and experiments are organized using an Author class. Each finger or array object can be associated with multiple Articles or Experiments. The objects representing the Java classes are mapped to a MySQL version 4.0.15 database using Hibernate.

Apache Tomcat version 5.5.15 is used as the servlet engine for the web site. The web application uses Struts 1.2 to implement the Model-View-Controller (MVC) pattern. Java Server Pages (JSP) are used to implement the user's view. The web application was developed using Eclipse 3.0.

CONCLUSIONS

ZiFDB is one of an increasing repertoire of tools for zinc-finger protein engineering provided by the Zinc Finger Consortium. ZiFDB houses information on individual zinc fingers, including the 141 zinc-finger modules currently made available as a clone library by the Consortium. Importantly, the database also provides users with information on ZFAs and their target sequences, and we believe this feature will be particularly

valuable for the rapidly growing number of molecular biologists interested in generating ZFAs for modification or regulation of their genome locus of interest. Further, analysis of the data housed in ZiFDB will also likely provide new insight into how zinc-finger proteins recognize their DNA targets, which, in turn, may lead to additional improvements in ZFA engineering.

Future perspective

Future versions of ZiFDB will allow users to directly input information on novel zinc fingers or ZFAs. To ensure data quality, a shadow database will be created that is associated with the persistent database. Users will be permitted to submit data to the shadow database where it will be held pending approval by the curator and subsequent loading into the persistent database. This feature will ensure that the database capitalizes upon new information generated by the scientific community, including unpublished ZFAs, and thereby should expand the utility of the database.

REFERENCES

1. Wolfe,S., Nekludova,L. and Pabo,C. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
2. Cathomen,T. and Joung,J.K. (2008) Zinc-finger nucleases: the next generation emerges. *Mol. Ther.*, **16**, 1200–1207.
3. Dreier,B., Beerli,R., Segal,D., Flippin,J. and Barbas,C.F. (2001) Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466–29478.
4. Dreier,B., Fuller,R., Segal,D., Lund,C., Blancafort,P., Huber,A., Koksch,B. and Barbas,C.F. (2005) Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **280**, 35588–35597.
5. Dreier,B., Segal,D. and Barbas,C.F. (2000) Insights into the molecular recognition of the 5′-GNN-3′ family of DNA sequences by zinc finger domains. *J. Mol. Biol.*, **303**, 489–502.
6. Segal,D., Dreier,B., Beerli,R. and Barbas,C.F. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
7. Liu,Q., Xia,Z., Zhong,X. and Case,C.C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.*, **277**, 3850–3856.
8. Desjarlais,J. and Berg,J.M. (1992) Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl Acad. Sci. USA*, **89**, 7345–7349.
9. Sera,T. and Uranga,C. (2002) Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry*, **41**, 7074–7081.
10. Bae,K., Kwon,Y., Shin,H., Hwang,M., Ryu,E., Park,K., Yang,H., Lee,D., Lee,Y., Park,J. *et al.* (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.*, **21**, 275–280.
11. Ramirez,C., Foley,J., Wright,D., Muller-Lerch,F., Rahman,S., Cornu,T., Winfrey,R., Sander,J., Fu,F., Townsend,J. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
12. Greisman,H. and Pabo,C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
13. Isalan,M., Choo,Y. and Klug,A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl Acad. Sci. USA*, **94**, 5617–5621.
14. Joung,J.K., Ramm,E. and Pabo,C.O. (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **97**, 7382–7387.
15. Hurt,J., Thibodeau,S., Hirsh,A., Pabo,C.O. and Joung,J.K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl Acad. Sci. USA*, **100**, 12271–12276.
16. Maeder,M., Thibodeau-Beganny,S., Osiak,A., Wright,D., Anthony,R., Eichtinger,M., Jiang,T., Foley,J., Winfrey,R., Townsend,J. *et al.* (2008) Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
17. Wright,D.A., Thibodeau-Beganny,S., Sander,J.D., Winfrey,R., Hirsh,A., Eichtinger,M., Fu,F., Porteus,M., Dobbs,D., Voytas,D. *et al.* (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat. Protoc.*, **1**, 1637–1652.
18. Sander,J., Zaback,P., Joung,J.K., Voytas,D. and Dobbs,D. (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res.*, **35**, W599–W605.
19. Zhang,L., Spratt,S., Liu,Q., Johnstone,B., Qi,H., Raschke,E., Jamieson,A., Rebar,E.J., Wolffe,A.P. and Case,C.C. (2000) Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *J. Biol. Chem.*, **275**, 33850–33860.
20. Liu,P., Rebar,E., Zhang,L., Liu,Q., Jamieson,A., Liang,Y., Qi,H., Li,P.X., Chen,B., Mendel,M.C. *et al.* (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.*, **276**, 11323–11334.
21. Mandell,J. and Barbas,C.F. (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.*, **34**, W516–W523.